A new regularity based descriptor computed from local image oscillations.

Leonardo Trujillo, Gustavo Olague, Pierrick Legrand, Evelyne Lutton

A new regularity based descriptor computed from local image oscillations

Leonardo Trujillo[1], **Gustavo Olague**[1*],
Pierrick Legrand[2] **and Evelyne Lutton**[2]

[1]*EvoVisión Project, CICESE Research Center, Applied Physics Division,*
Km. 107 carretera Tijuana-Ensenada 22680, Ensenada, B.C. México

[2]*Complex Team, INRIA Roquencourt,*
Domaine de Voluceau, BP 105 78153 Le Chesnay Cedex, France

[*]*olague@cicese.mx*

Abstract: This work presents a novel local image descriptor based on the concept of pointwise signal regularity. Local image regions are extracted using either an interest point or an interest region detector, and discriminative feature vectors are constructed by uniformly sampling the pointwise Hölderian regularity around each region center. Regularity estimation is performed using local image oscillations, the most straightforward method directly derived from the definition of the Hölder exponent. Furthermore, estimating the Hölder exponent in this manner has proven to be superior when compared to wavelet based estimation. Our detector shows invariance to illumination change, JPEG compression, image rotation and scale change. Results show that the proposed descriptor is stable with respect to variations in imaging conditions, and reliable performance metrics prove it to be comparable and in some instances better than SIFT, the state-of-the-art in local descriptors.

OCIS codes: (000.0000) General.

References and links

1. C.J.G. Evertsz and B.B. Mandelbrot. Multifractal Measures. In *Chaos and Fractals: New Frontiers in Science*, pp. 849–881. H.-O. Peitgen, H. Jurgens and D. Saupe, Springer, New-York, 1992.
2. K. Falconer. Fractal geometry, Mathematical Foundations and Applications, Wiley, Chichester, 1990.
3. P. Legrand, E. Lutton and G. Olague. Evolutionary denoising based on an estimation of H¨older exponents with oscillations. In *EVOIASP Workshop*, pp. 520-524, Budapest, 2006.
4. P. Legrand and J. Lévy Véhel. Local regularity-based interpolation. In *WAVELET X*, Part of SPIE's Symposium on Optical Science and Technology, San Diego, CA, August 3-8, 2003, proceedings of SPIE Vol. 5207.
5. P. Legrand. Debruitage et interpolation par analyse de la regularite H¨olderienne. Application a la modelisation du frottement pneumatique-chaussee. PhD thesis, Université de Nantes.
6. J. Lévy Véhel. Fractal Approaches in Signal Processing. *Fractals*, 3(4):755–775, 1995.
7. D.G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *Intl J. Computer Vision*, 2(60):91–110, 2004.
8. K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *Intl J. Computer Vision*, 1(60):63–86, 2004.
9. H. P. Moravec. Towards automatic visual obstacle avoidance. In *IJCAI*, page 584, 1977.
10. C. Schmid and K. Mikolajczyk. A Performance Evaluation of Local Descriptors. *Pattern Recognition and Machine Intelligence*, 27(10):1615–1630, 2005.
11. C. Tricot. Curves and Fractal Dimension. *Springer-Verlag*, 1995.
12. L. Trujillo and G. Olague. Synthesis of interest point detectors through genetic programming. In *GECCO 2006*, volume 1, pages 887–894, Sattle WA, USA, July 2006. ACM Press.

13. L. Trujillo and G. Olague. Using Evolution to Learn How to Perform Interest Point Detection. In *Proceedings of the 18th International Conference on Pattern Recognition*, volume 1, pages 211–214, Hong Kong, August 2006.
14. http://www.robots.ox.ac.uk/ vgg/research/

---

1. Introduction

The feature extraction problem, in the domain of image analysis systems, poses to main research questions. How can *distinctive* areas within an image be **identified**? How can distinctive areas be **represented** in such a way as to facilitate their identification? The main concepts to be taken from those questions are: identification and representation. Concerning the former, the identification problem, a mainstay in vision systems are interest point or interest region extraction algorithms. These techniques search for image pixels, or image regions, that exhibit high signal variations with respect to a particular local measure. Solutions have been designed based on studying intrinsic properties of 2D-signals [9], and more recently by solving a properly framed optimization problem [12, 13]. In response to the second question, dealing with the concept of representation, different techniques hanve been proposed that encode the information within these so called interesting regions. Discriminative feature are constructed that uniquiley characterize each interest regions. This in turn allows for efficient feature matching in a wide range of imaging problems. Currently, the SIFT [7] descriptor has proven to be the most discriminative local descriptor in machine vision literature, and shows the highest performance with respect to the current set of benchmark tests [10].

This paper presents a novel region descriptor based on the concept of Hölderian regularity. By approximating the pointwise Hölder exponent, also known as the Lipschitz exponent, using local signal oscillations around each image point, we are able to construct discriminative feature vectors. Our proposed descriptor is invariant to several types of changes in viewing conditions, exhibiting high and stable performace. As such, the main contribution of this work is that it introduces novel concepts to the field of feature extraction algorithms, using formal mathematical tools and corroborated by high performance on standard tests.

The rest of this paper is organized as follows. Section 2 gives a brief overview of related work. Section 3 presents the concept of Hölderian regularity and how to estimate it. Section 4 introduces our local descriptor based on pointwise Hölder exponents. Later in Section 5, experimental results are provided. Finally, in Section 6 we give conclusions and outline possible future work.

2. Related Work

It is not our intention to give a comprehensive summary on the subject of local descriptors, such a discussion can be found in [10]. Hence, we will only focus on presenting the basic strategies followed by the most common type of region detectors, distribution based descriptors, and discuss the SIFT strategy.

Currently, most state-of-the-art local descriptors use a distribution based approach. These techniques characterize image information using local histograms of a particular measure related to shape or appearance. The most simple would be using histograms of pixel values, while more complex representations could be values representing texture characteristics. The most successful descriptor currently available in computer vision literature is SIFT, developed by David Lowe [7], which builds an histogram of gradient distributions within an interest region. The descriptor builds a 3D histogram of gradient locations and orientations, weighted by the gradient magnitudes. Although SIFT combines both a scale invariant detector with the gradient distribution descriptor, only the latter has proven to outperform other types of techniques, and it is possible to replace the former with a more reliable region detector.

3. Hölder Regularity

One of the most popular ways to measure a signals regularity, be it pointwise or local, is to consider Hölder spaces. Hence, we will present the concept of regularity expressed through the Hölder exponent.

DEFINITION 1. Let $f : \Re \to \Re$, $s \in \Re^{+*} \setminus \mathbf{N}$ and $x_0 \in \Re$. Then, $f \in C^s(x_0) \Leftrightarrow \exists \eta \in \Re^{+*}$, a polynum P of degree $< s$ and a constant c such that

$$\forall x \in B(x_0, \eta), |f(x) - P(x - x_0)| \le c|x - x_0|^s. \tag{1}$$

The pointwise Hölder exponent of f at x_0 is $\alpha_p = \sup_s \{f \in C^s(x_0)\}$, see Figure 1.

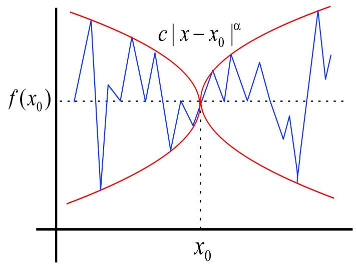

Fig. 1. H¨olderian envelope of signal f at point x_0.

The concept of signal regularity, characterized by the Hölder exponent, has been widely used in fractal analysis [2, 1]. With regards to image analysis, the Hölder exponent provides a great deal of information related to the local structure around each point. Hence, it has been applied to such tasks as edge detection [6], image denoising [3] and image interpolation [4]. Furthermore, because most local image descriptors are fundamentally attempting to describe local image variations and overall structure, it is a natural conclusion to expect that Hölderian regularity will prove to be a useful tool in this task. Now, we are left with the task of accurately estimating the pointwise Hölder exponent.

3.1. Estimating the Hölder Exponent with oscillations

The most natural way to estimate the Hölder exponent, because it follows from its definition, consists in studying the oscillations around each point. This method gives accurate results, better then those obtained using wavelet analysis [5], hence it will be the technique of choice to compute our proposed descriptor. A brief description of this technique will now be given, for a more detailed analysis please see [11]. It is pointed out that the Hölder exponent of function $f(t)$ at t is $\alpha_p \in [0, 1]$, if a constant c exists such that $\forall\, t'$ in a vicinity of t,

$$|f(t) - f(t')| \le c|t - t'|^{\alpha}_p. \tag{2}$$

In terms of signal oscillations, this condition can be written as: a function $f(t)$ is Hölderian with exponent $\alpha_p \in [0, 1]$ at t if $\exists c\, \forall \tau$ such that $osc_\tau(t) \le c\tau^{\alpha_p}$, with

$$osc_\tau(t) = \sup_{|t-t'|\le\tau} f(t') - \inf_{|t-t'|\le\tau} f(t') = \sup_{t',t''\in[t-\tau,t+\tau]} |f(t') - f(t'')|. \tag{3}$$

An estimation of the regularity will be built at each point by computing the slope of the regression between the logarithm of the oscillation and the logarithm of the dimension of the

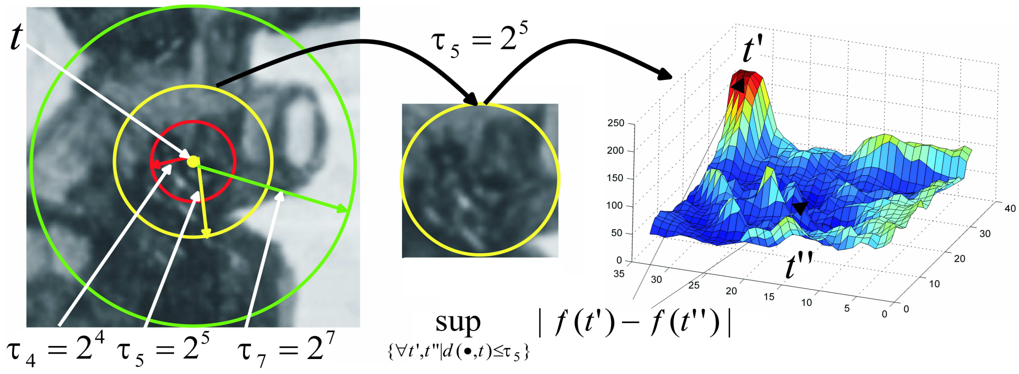

Fig. 2. Estimating the H¨older exponent with oscillations. **Left**: the region of interest λ, and three of the seven neighborhoods around point t, when $r = 1,2,\cdots,7$. **Center**: the neighborhood of radius $\tau_5 = 32$ pixels, with $base = 2$' **Right**: computing the supremum of the differences within radius τ_5, where d denotes the Euclidian distance.

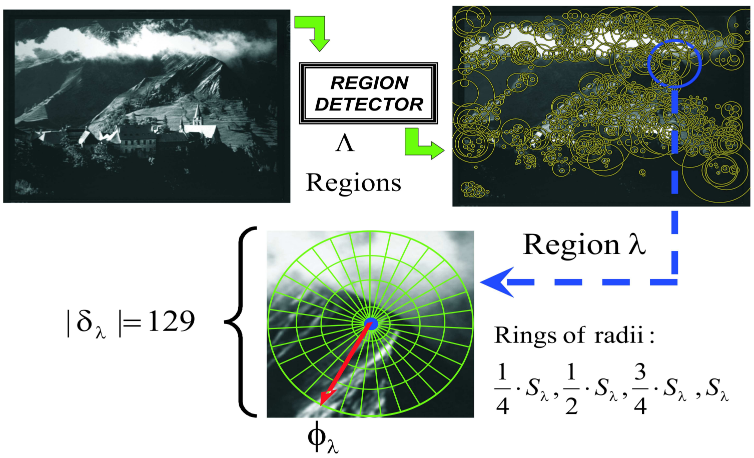

Fig. 3. Descriptor building process.

neighborhood at which one calculates the oscillation. From an algorithmic point of view, it is preferable not to use all sizes of neighborhoods between two values τ_{min} and τ_{max}. Hence, we calculate the oscillation at point t only on intervals of the form $[t - \tau_r : t + \tau_r]$, where $\tau_r = base^r$. Here, we use least squares regression, with $base = 2$ and $r = 1,2,\ldots,7$. For a 2D signal, t defines a point in 2D space and τ_r a radius around t, such that the Euclidean distances $d(t',t)$ and $d(t'',t)$ are $\leq \tau_r$. We can visualize this process in Figure 2. The method of estimation with oscillations will give good results under three conditions: that $\alpha_p < 1$, the regression converges, and the regression converges towards a valid slope.

4. Hölder Descriptor

Now that we have described a method to accurately characterize the pointwise signal regularity, we can now move on to describe how we use this information to build our local descriptors. The process, described in Figure 3, is as follows.

First, a set Λ of regions of interest are extracted from an image. Second, the dominant gradient orientation ϕ_λ is computed, this preserves rotation invariance. Finally, our feature vector δ_λ contains the Holder exponent α_p of the region center and of 128 concentric points, orderd according to ϕ_λ.

Region Extraction: The first step in the process requires stable detection of prominent image regions. The type of regions to be extracted will depend on the requirement of the higher level application with respect to invariance. For instance, an interest point detector will suffice when the scale of the imaged scene is not modified. In our work, we use a detector optimized for geometric stability and global point separability, the IPGP2 detector which is the determinant of the Hessian matrix smoothed by a 2D Gaussian [12, 13]. All regions extracted with an interest point detector are assigned the same scale, $w_\lambda = 2.5$ pixels. For images where scale is a factor, we use the Hessian-Laplace detector presented in [8], which searches for extrema in a linear scale space generated with a Gaussian kernel. After this step we are left with a set Λ of circular image regions, where the scale is set to $s_\lambda = 5 \cdot w_\lambda$, and w_λ is the scale given by the detector.

Dominant Orientation: In order to preserve rotation invariance, the dominant gradient orientation is computed and used as a reference for the subsequent sampling process. For the scale invariant detector, all image regions are normalized to 41x41 bit size using bicubic interpolation. An orientation histogram is constructed using gradient orientations within the interest region, similar to what is described in [7]. The histogram peak is obtained and thus $\forall \lambda \in \Lambda$ a corresponding ϕ_λ is assigned. In this way, each region is described by a set $\lambda = \{x_\lambda, y_\lambda, s_\lambda, \phi_\lambda\}$, the image center, scale and orientation of the region.

Hölder Descriptor: Now that regions are appropriately detected and described with λ, we can now continue to construct our region descriptor $\delta_\lambda, \forall \lambda \in \Lambda$. Our sampling process is simple, see Figure 3, the first element of δ_λ is the Hölder exponent α_p computed at the region center (x_λ, y_λ). Next, the Hölder exponent of points on the perimeter of four concentric rings are sampled, with radii of $\frac{1}{4} \cdot s_\lambda$, $\frac{1}{2} \cdot s_\lambda$, $\frac{3}{4} \cdot s_\lambda$ and s_λ respectively. A total of 32 points on each ring are sampled, starting from the position given by ϕ_λ, uniformly spaced and ordered counterclockwise. Hence, our feature vector δ_λ has 129 dimensions, compared to the 128 of SIFT.

5. Experimental Results

In order to effectively evaluate and compare our results, we use standard image sequences provided by the Visual Geometry Group [14]. From each image sequence there is one reference image and a set of test images, since we know beforehand the transformation between the reference and test images we are able to quantify a matching score for our descriptor. For image sequences where there is no scale change, we use threshold based matching, and for images with scale change we use neareast neighbor distance ratio matching. The former, is a strategy where two image regions λ_1 and λ_2 are matched if the following relation holds $d(\delta_{\lambda_1}, \delta_{\lambda_2}) < t$. While the latter strategy assigns a match between regions if $\frac{d(\delta_{\lambda_1}, \delta_{\lambda_2})}{d(\delta_{\lambda_1}, \delta_{\lambda_3})} < t$, where λ_2 is the nearest neighbor of λ_1, and λ_3 is the second nearest. In both cases, the value of t is varied to obtain the performance curves. Two types of curves are presented: one plots *recall* versus *1-precision*, characterizing the matching between one test image and the reference image [10]; the other is a double *y-axis* plot, one axis for *recall* and the other for *1-precision*, that characterizes the performance of the descriptor on an entire image sequence. The second type of plot, includes errorbars in order to visualize the stability of the descriptor. Recall and 1-precision are defined as in [10]: $recall = \frac{\#correctmatches}{\#correspondences}$, and $1 - precision = \frac{\#falsematches}{\#correctmatches - \#falsematches}$. For comparison, the performance of our descriptor is plotted with that of SIFT. To compute SIFT descriptor, the Harris and Harris-Lapplace detectors were used to extract image regions, as suggested in [10]; executables for SIFT and the Harris detectors were obtained from [14].

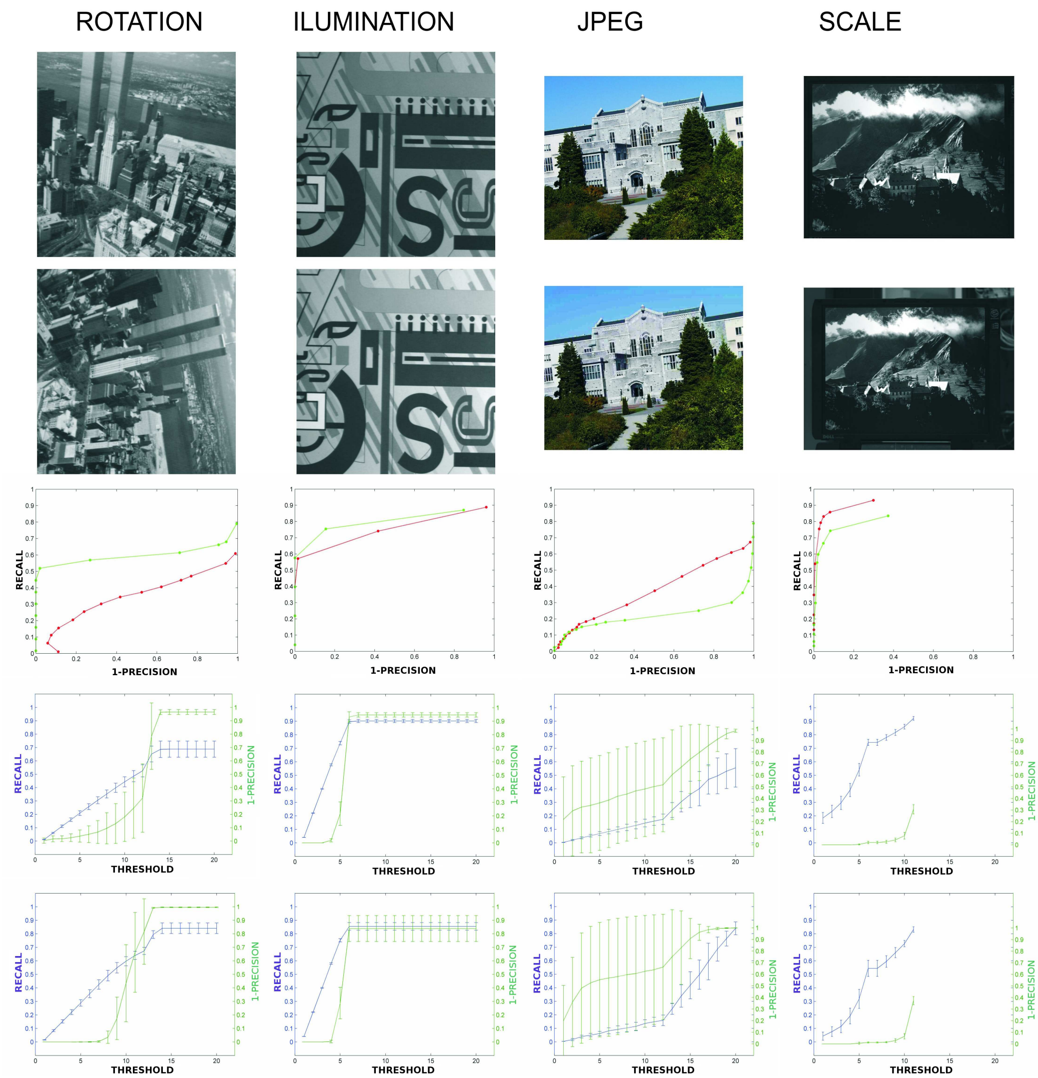

Fig. 4. **Columns**, from left to right: 1)Rotation (36 images in sequence), 2)Illumination change (10 images), 3)JPEG compression (6 images), and 4)Scale change (first 6 images of sequence). **Rows**, from top to bottom: 1)Reference image, 2)Test Image, 3)Performace between test and reference with Hölder-Green and SIFT-Red, 4)SIFT average performance on entire set, and 5)Hölder average performance.

6. Conclusions and Future Work

Results show very promising performance, in general we can appreciate how the regularity based descriptor is more stable and achieves equal or better performace than SIFT for image sequences without scale change. Even do this is not the case for scale change transformations, we can still appreciate competitive performace up to a reasonable change in scale. The performance drop-off in this circumstances is expected to be directly related to the method of Hölder exponent estimation. For this reason an appropriate modification of the oscillations method is necessary in order to obtain a more efficient scale invariance for our Hölder descriptor.